TRAVEL TIME DATA QUALITY ASSESSMENTS

STATE OF THE PRACTICE

OVERVIEW

A number of data quality assessments of probe-based travel time estimation technology and data service providers have been recently conducted. In some cases these evaluations were conducted by universities, in other cases by third party consultants. The goal in each case was to measure the accuracy of speed and travel time estimates coming from a travel time data service provider or probe-based technology. To measure the accuracy of the data, an attempt is made to collect ground truth data over some spatial and temporal extent and compared to the estimate. The accuracy of the data is measured by a performance metric such as MAE (mean absolute error), Error Bias or RMSE (root mean squared error). However, in order to measure an estimate's error it is necessary to agree upon how to measure the ground truth. There have been some notable differences in the methods used to collect ground truth data and in the statistical methods used to measure accuracy. This document provides an overview of past efforts to measure the accuracy of probe based travel time estimation technologies including the selected ground truth data collection technologies and the error metrics used.

GROUND TRUTH DATA COLLECTION TECHNOLOGIES

Ground truth data can be defined as any observations of travel time over some spatial and temporal extent that is used as reference data for evaluating some travel time estimation method. There are a number of different ways ground truth data can be collected. The differences in how ground truth is measured can affect the resulting findings of the data quality assessment. The major technologies used in past evaluations are point sensors, like inductive loop detectors, or probe based methods, like floating car runs or Bluetooth data samples. This section will provide a brief overview of the most commonly used technologies for collecting ground truth data and provide some background on the advantages and disadvantages of each technology.

POINT SENSORS

A point sensor measures the presence and speed of every vehicle that travels by the "point" where it is deployed or aimed. However, there are some problems when using point sensor data to measure travel times. Speed estimates derived from point sensor data are point estimates of speed averaged over time. This is often referred to as the *time mean speed*. Travel time, on the other hand, is the time it takes to travel from one point on a network to any other point on the network. The time and distance traveled can be used to calculate the average speed for that trip. This is often referred to as the *space mean speed*. A time mean speed observation at any point on a segment only gives an observation of the average speed of vehicles traveling over that particular point during some time interval. It does not necessarily reflect the conditions upstream or downstream of the point. Space mean speed can be estimated from time mean speed measurements. However the accuracy of travel time estimates from point sensor data tends to decrease as congestion levels increase. In general, point sensor data works well for validating ground truth measurements from other data sources, but should not be used as the sole source of ground truth data for assessments of travel time data.

PROBE DATA SYSTEMS

Probe-based data collection technologies such as floating car data (FCD) and automated vehicle identification (AVI) technologies can be used to estimate travel time by measuring travel time from samples of vehicles in the traffic stream. This approach to measuring ground truth has the advantage over loop detector data in that it directly measures travel times rather than estimating travel times from point sensors. Additionally, statistical tests can be applied to the sample data to provide a degree of confidence in the accuracy of the ground truth data.

Statistical Sampling And Estimation

To measure ground truth data with statistical confidence requires a sample set of observations. A sample set is any number of randomly selected observations from a specified population. The population of interest in this case is the set of all vehicles having traveled a specific segment during some specific time period. If we could measure the travel time of every vehicle traversing that segment we could calculate the average travel time exactly of the population. However, by taking just a sample of observations from the population we can estimate the population mean by calculating the sample mean and standard deviation. The number of samples needed depends on the variance and desired precision of the estimate. The equation for the minimum sample size is:

$$n = \left(\frac{z_{\alpha/2}\frac{\sigma}{\mu}}{e}\right)^2$$

where $z_{\alpha/2}$ is the critical normal deviate usually given as 1.65 for 90% degree of confidence, $\frac{\sigma}{\mu}$ is the standard deviation over the mean (termed the coefficient of variation of the sample data), and e is the error tolerance level given as a percentage value (i.e. 10% = 0.10).

For example, if we have a sample of 10 observations with the average space mean speed equal to 55 mph and the standard deviation equal to 10 mph, we can compute that the minimum sample size needed to estimate the average speed for all vehicles traversing that link at the time of observation. If a 10 percent error tolerance is considered acceptable, the minimum number of samples required to estimate the mean to this level of accuracy would be:

$$n = \left(\frac{1.65 * \frac{10}{55}}{0.10}\right)^2 = 9$$

Since we observed 10 vehicles and the minimum sample size was computed to be 9 vehicles, we can say with 90% degree of confidence that the average speed of the population is 55 mph, +/- 10%.

Floating Car Data

Floating car data derived from GPS time/location points can directly measure travel times by mapping the GPS data points to a segment and determining the time taken by the vehicle to travel over the segment. Thus, floating car data provides a measurement of travel time and space mean speed for the segment at a particular time. Floating car data has been widely used as ground truth because it directly measures the travel time over the segment and is relatively simple to collect. However, each floating car measurement provides only a single observation and may not represent the average travel time experienced by all vehicles over the same segment during the same period of time. For example, consider a vehicle traveling in heavily congested flows. The travel times experienced in these conditions are not uniform. In some cases, the floating car observation can be higher or lower than the average of the population. In many cases floating car data does a good job of estimating the mean of the distribution, but there may also be conditions where greater statistical confidence in the ground truth data is needed.

Automatic Vehicle Identification (AVI) Technologies

Automatic Vehicle Identification consists of technologies that can be used to identify and re-identify specific vehicles from a traffic stream traveling over a defined spatial and temporal extent. Typically these technologies capture a unique identifier from the vehicle and use this identifier to re-identify the vehicle at a point further down the road. The two technologies most commonly used are Bluetooth readers and toll tag readers.

BLUETOOTH READERS

Bluetooth readers have been tested extensively by the University of Maryland’s Center for Advanced Transportation Technology (CATT) in their evaluation of Inrix’s data quality for the I-95 Corridor Coalition. The University of Virginia, Purdue University, the University of Washington, and the Texas Transportation Institute, among others, are also in the process of developing Bluetooth readers. The technology works by detecting Bluetooth data communication in a vehicle and reading the MAC address in the data signal that can be used to re-identify the vehicle further down the road. One of the advantages of this technology is that it can be rapidly deployed to different sites by placing a reader at each end of the segment where travel time measurements are to be recorded. The technology is also relatively inexpensive.

TOLL TAG READERS

Toll tag readers are another common type of AVI sensor. Toll tag data collection works by identifying vehicles based on radio frequency identification (RFID) tags deployed in vehicles for the purpose of automated toll collection on tolled facilities. The tags can be uniquely identified at points on the road where readers have been deployed. For example, an extensive toll tag infrastructure has been deployed in the Houston metropolitan area and is being used to collect large volumes of travel time data on the roads in the network. The limitation of this technology is that it depends on both the number of vehicles carrying the RFID tags (penetration rate) and the extent of the deployed reader infrastructure.

MEASURING ERROR

The difference between a ground truth measurement of travel time and an estimate from a service provider is the error in the service provider’s estimate. There are a number of different ways to measure error and the choice of an error metric can affect the results of a quality evaluation. One of the most commonly used error metric is the **Mean Absolute Error (MAE)** which is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} abs(T_i - \hat{T}_i)$$

where T is the ground truth travel time, $\hat{T}$ is the service provider’s estimate of travel time, and n is the number of observations to be evaluated.

The MAE is the average of the absolute value of differences between the ground truth measurement and the service provider’s estimate. It gives a measure of the average magnitude of error in a service provider’s data. However, the MAE does not indicate whether the estimates tend to be over-estimates or under-estimates.

A second metric that is used is the **Error Bias** which is the average of all errors without taking the absolute value of each error. A positive error bias will indicate that the service provider tends to under-estimate travel times. Similarly a negative error bias will indicate that the service provider tends to over-estimate travel times. However, both MAE and Error Bias are not as sensitive to large errors (i.e. outlier estimates) as other metrics. This can be a problem when it is important to identify cases where a service provider has many accurate estimates but also has a few estimates that are particularly far off from the ground truth. The Root Mean Square Error (RMSE) can help identify these cases by squaring the errors first and then taking an average of the squared errors and finally taking the square root of the average to report the metric in the base units. Because squaring is a non-linear operation it weights outlier observations more heavily and gives a better indication of whether a data set contains outlier observations. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t}^{n}\left(T_t - \hat{T}_t\right)^2}$$

Other error metrics commonly used include the MAPE (mean absolute percent error) and Percent Error Bias. These errors report the relative deviations (in percentages) between ground truth and the service provider's estimates.

The choice of an error metric depends on the goal of the evaluation. Error metrics such as MAE, Error Bias, and RMSE can quantify the error in a service provider's data. However, an evaluation may also use error ranges or "bins" to categorize the degree of error. Several different error metrics can be used and a decision can be based on the results of all metrics.

SELECTED DATA QUALITY ASSESSMENTS

While private sector probe data quality evaluations have been conducted as far back as 1994 for the CAPITAL test in the Washington, DC region, this document will highlight some of the more recent evaluations. The three major data service providers that have been evaluated in the United States are Inrix, Airsage, and Cellint. Both Airsage and Cellint are data service providers that estimate travel time from cellular phone data. Inrix estimates travel time from a fusion of commercial GPS data, DOT sensor data, and other proprietary data sources. In addition to evaluations of these three vendors, there have been a number of other efforts to develop and/or evaluate travel time data technologies. For example, the Mobile Millennium project at the University of California, Berkeley focused on evaluating the use of Smart Phone technology to estimate travel times. Research at the University of Akron focused on an evaluation of data posted on variable message signs and NAVTEQ conducts an ongoing audit of traveler information in a number of different markets.

The results of these evaluations have been mixed. The evaluation of Inrix by the University of Maryland and the I-95 corridor coalition has been extensive and the results are publicly available through the website (see below). One of the important aspects of this evaluation is the use of Bluetooth data readers for measuring ground truth. Most of the other evaluations have used either floating car or loop detector measurements as ground truth. This is understandable given the relatively recent development of Bluetooth reader technology. However, it may be the case that future evaluations will be using a mix of AVI and floating car data for evaluations.

Table 1 shows a summary of the selected data quality assessments reviewed in this document. The table shows the time frame for when the evaluation was conducted, a list of ground truth technologies used in the evaluation, a list of relevant error metrics used, and a synopsis of the study results. Detailed descriptions of each study follow the table.

TABLE 1 - SUMMARY OF SELECTED EVALUATIONS

Service Provider	Evaluator	Time Frame	Ground Truth Data	Error Metric	Results
Inrix	University of Maryland	On-going since 2008	Bluetooth (with floating car validation)	MAE, Speed Error Bias	Study shows that speed estimates are meeting the terms of the contract (i.e. < 10 mph error)
Inrix, Traffic.com	Frost & Sullivan	2006	Floating car	RMSE	Frost & Sullivan reported that Inrix was the “market leader” (their words) in traveler information systems.
AirSage	University of Virginia	2005	Floating car data and loop detectors	MAE, Speed Error Bias	68% of speed estimates had greater than 20 mph error
AirSage	GeoStats	2008	Floating car data	% of times of congestion detected	Three markets tested, found > 85% of time congestion correctly detected
Cellint	URS, GeoStats, Georgia DOT	2007	Floating Car and calibrated loop detector model	Paired t-test of means	Significant match in speeds between 20 and 70 mph, below 20 mph did not perform as well
Mobile Millennium Project	University of California	2008	Loop detectors	Absolute percent error	Demonstrated that less than 5% penetration rate of probes could provide accurate estimates of speeds

SUMMARIES OF DATA QUALITY ASSESSMENTS

INRIX

UNIVERSITY OF MARYLAND / I-95 CORRIDOR COALITION EVALUATION (1)

PROJECT SUMMARY

This is an ongoing effort to evaluate the quality of travel time data being provided by Inrix to the I-95 Corridor Coalition. The ground truth data used in this evaluation is measured by Bluetooth readers that can identify vehicles based on the MAC address present in all Bluetooth communication. Over 100 miles of highways across Delaware, Maryland, New Jersey and Virginia were examined.

METHODOLOGY

Speeds were computed from the ground truth travel times measured by the Bluetooth readers. These speeds were compared to the corresponding estimated speed provided by Inrix. The absolute error and error bias were measured for each segment and 5-minute time period. The speeds were grouped into bins (0-30 mph, 30-45 mph, 45-60 mph, 60+ mph) and the average absolute error (MAE) and speed error bias (average error including sign) were computed for each bin in each study area. More information on the study can be found at the following URL.

http://www.i95coalition.net/i95/Projects/ProjectDatabase/tabid/120/agentType/View/PropertyID/107/Default.aspx

FROST & SULLIVAN EVALUATION OF INRIX AND TRAFFIC.COM (2)

PROJECT SUMMARY

This study was an evaluation of Inrix and Traffic.com in three different markets: Philadelphia, PA; Providence, RI; and Washington, DC. The service providers were evaluated on the basis of accuracy and breadth of coverage in the market. The Traffic.com data is mostly derived from loop detectors whereas the Inrix data is fused data from GPS enabled commercial fleets and loop detectors.

METHODOLOGY

Around 141 floating car runs were used to measure accuracy in the three markets. Root-mean-square-error (RMSE) was used to measure the accuracy of the Inrix data when compared with ground truth. Overall the results were positive for both providers with Inrix getting a slightly better evaluation.

CELLINT

URS & GEOSTATS / GEORGIA DOT EVALUATION IN THE ATLANTA METROPOLITAN REGION (3)

PROJECT SUMMARY

URS & GeoStats were contracted by Georgia DOT to evaluate real-time traveler information originating from CellInt, a data service provider, and from NaviGAtor which is a system developed in Georgia that uses loop detector data to estimate travel times. The Cellint data was deemed to be acceptably accurate in this study. The report notes that inadequate sample size in the ground truth data was a problem in the evaluation.

METHODOLOGY

Cellint data was compared with data from calibrated loop detectors. The loop detectors were calibrated by using floating car data. The objective of the study was to determine the accuracy of the Cellint data compared with the data from calibrated loop detectors and from the floating car runs.

KANSAS DOT / MISSOURI DOT EVALUATION (4)

PROJECT SUMMARY

Kansas DOT and Missouri DOT conducted an evaluation of CellInt TrafficSense system. Only speeds were evaluated. The test was conducted on I-435 in Kansas City from I-470 to I-35.

METHODOLOGY

Loop detector data was collected and used as ground truth. The system was tested for latency and accuracy. Overall the report indicated that the system met the expected accuracy requirements while latency of 5-7 minutes was considered to be somewhat problematic.

AIRSAGE

HAMPTON ROADS AND NORTHERN VIRGINIA EVALUATIONS (5) (6)

PROJECT SUMMARY

University of Virginia evaluated AirSage system in Hampton Roads and in Northern Virginia in 2005 and 2008, respectively. The Hampton Roads results showed 68% of estimates with an error greater than 20 mph. The Northern Virginia evaluation found that 83% of possible five minute intervals had a valid speed/travel time estimate and that Airsage estimates differed from the baseline link speed by 7 to 13 mph, on average.

METHODOLOGY

The Hampton Roads evaluation used floating car data to measure ground truth. The Northern Virginia evaluation used synchronized floating cars that entered the traffic stream to try to obtain an estimate of population variability in the traffic stream at a point in time. Although all probes entered at the same time, they were told to operate independently of each other. All of the Northern Virginia travel time runs were conducted during congested conditions.

GEOSTATS EVALUATION OF AIRSAGE (7)

PROJECT SUMMARY

Airsage data was evaluated in San Diego, New York, and Detroit by GeoStats, a third party evaluator. The goal of the evaluation was to measure the overall quality of Airsage's real-time traffic data. On average Airsage was able to correctly detect congestion about 85-90% of the times it occurred.

METHODOLOGY

Collected ground truth data using GPS floating car runs in October 2008. Probe vehicle drivers were dispatched at 15-minute intervals along the same route. Runs were conducted primarily during congested conditions and on both highways and arterials in the three markets. In each market several hundred TMC segments were tested resulting in several thousand centerline miles of roads tested in the three markets.

OTHER PROJECTS OF INTEREST

MOBILE MILLENNIUM / UNIVERSITY OF CALIFORNIA, BERKELEY (8)

Conducted evaluation of travel time data generated by SmartPhone applications acting as probes. This is a joint project with Nokia and University of California, Berkeley. The evaluation used floating car runs to provide ground truth data. Results were positive. Also, they are working with NATWG (North American Traffic Working Group) to develop a benchmark for evaluating quality of traffic information.

OHIO DOT

University of Akron and TTI evaluated the accuracy of travel times posted on Ohio DOT changeable message signs. The travel times were estimated from SpeedInfo fixed-point sensors, and the Ohio DOT's data service contract with SpeedInfo has a data accuracy performance specification. The Ohio DOT evaluation used both test vehicle runs and Bluetooth readers to establish reference travel times on freeways and signalized highways.

NAVTEQ INTERNAL QUALITY CONTROL TESTING

NAVTEQ does extensive quality testing on a routine basis in its traffic information markets using test vehicle runs. The results of these tests are used to confirm quality for current and potential traffic data customers. NAVTEQ has provided summaries of these data quality procedures and results to TTI; however, they are protected under a non-disclosure agreement.

LESSONS LEARNED

While there is currently no consensus on the best methods for measuring travel time data accuracy there are a few lessons that can be learned from the studies reviewed in this document.

- Travel Time is a Distribution
  - A number of these studies have relied on point estimates of travel time to measure accuracy. However, as variations in speeds increase, a single ground truth measurement may not be sufficient.
- Concentrate on measuring accuracy during transition states
  - Variations in travel time are much greater during the transition from free flow to congested traffic. More resources should be directed towards measuring accuracy during transition periods between the two states.
- Route Level analysis may be needed

    - Focusing entirely on link level analysis may not be sufficient. Many users of travel time data will be interested in the travel time for a route. Knowing the accuracy of data at the link level may not be sufficient to measure accuracy at the route level; particularly when a route is composed of many links.
- Measuring accuracy on signalized arterials will be more challenging
    - Accuracy on highways does not necessarily translate into accuracy on signalized arterials. Travel time estimates on arterials will prove to be much more challenging to data service providers and is an area that requires particular attention.

REFERENCES

1. **Haghani, Ali.** *I-95 Corridor Coalition Vehicle Probe Project: Validation of Inrix Data July-September 2008.* College Park, MD : I-95 Corridor Coalition, 2009.

2. **Frost & Sullivan.** *Real-Time Traffic Flow Ground Truth Testing Methodology Validation and Accuracy Measurement.* 2006.

3. **URS & GeoStats.** *Data Evaluation for CellInt Cellular Probe on Freeways.* Atlanta, GA : s.n., 2007.

4. **Telvent Farradyne, Inc.** *Kansas City Scout Evaluation of Cellular Probe Data.* Kansas City, KS : s.n., 2007.

5. **University of Virginia Center for Transportation Studies.** *Evaluation of an Early Generation Travel Time Data Service - Northern Virginia.* Charlottesville, VA : s.n., 2008.

6. —. *Wireless Location Technology-Based Traffic Monitoring Demonstration and Evaluation Project.* Charlottesville, VA : s.n., 2006.

7. **GeoStats.** *AirSage Traffic Data Quality - Final Report.* Atlanta, GA : s.n., 2009.

8. *Mobile Century Using GPS Mobile Phones as Traffic Sensors: A Field Experiment.* **Herrera, J. et. al.** New York : s.n., 2008. 15th World Congress on Intelligent Transportation Systems.